ilifu Data Deposit Guidelines

Contents

1. Before depositing your data

The problem with data in today's world is that there is too much of it. Too much to wade through to get to the data that might be useful to you. We are producing data at an exponential rate, and we disagree about which data are useful and which are not. The state of the world's data can best be described by this little diagram.
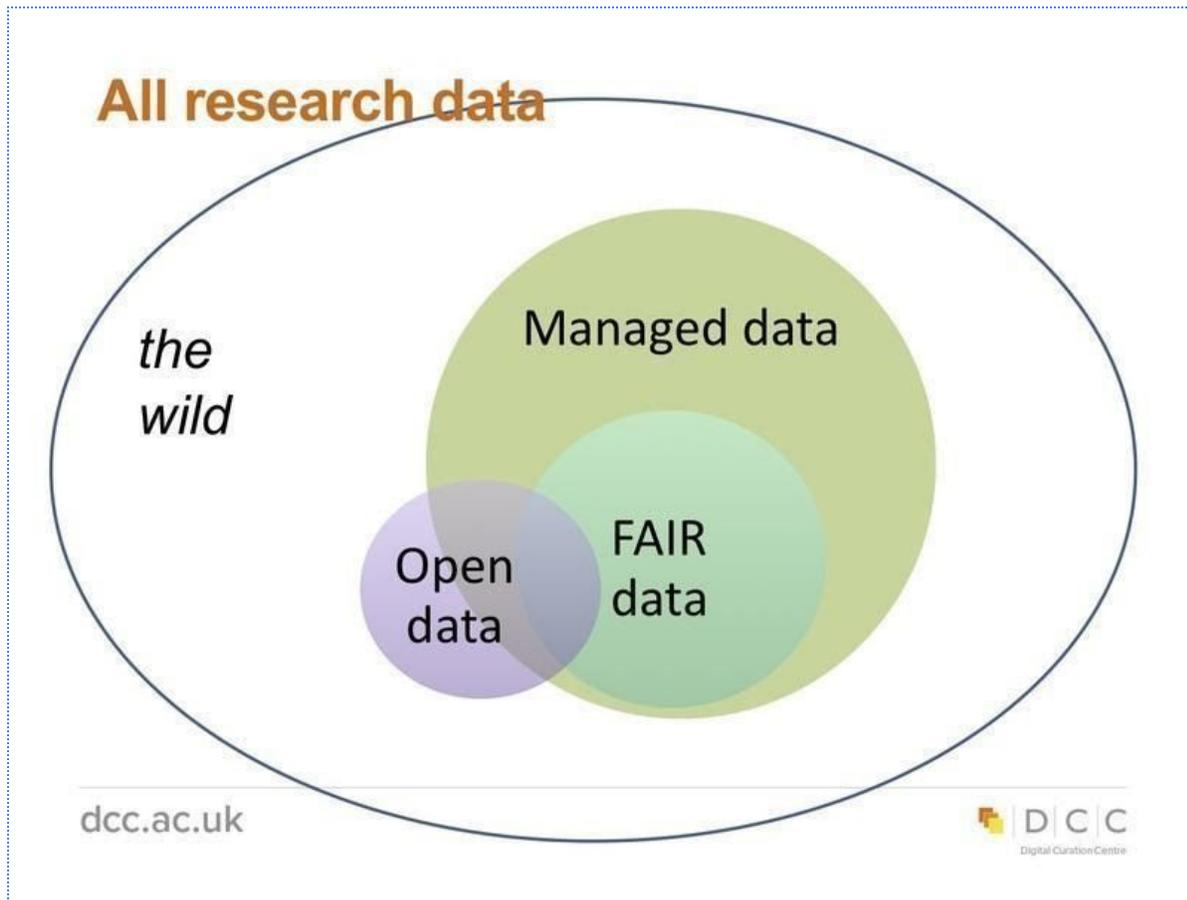

As a researcher your research data are valuable and most probably very useful. By depositing, and so sharing or publishing them, you want the rest of the world to be able to interact with the data sets that you have generated. However, the data by themselves are not enough. How will other researchers know where to find them? How will they know how to reuse them? How will they know if the data are relevant to their research? All of these questions need to be answered when you deposit the data in order to ensure that the context associated with the data is described. This will serve to assist others find and eventually reuse the data.

This is why it is important to consider the FAIR data principles when depositing any research data.These principles advocate for research data to be findable, accessible, interoperable and reusable.[1] The table below provides some more details on the nature of each of these principles.

Findable	...ensuring that your data can be found by both humans and machines.
Accessible	...once someone has found your data, they need to know how they can get access to them. This could include going through an authorisation and/or authentication process (doesn't have to be open).
Interoperable	...ensuring that your data can be integrated with other data and that they can be utilised by applications or workflows for analysis, storage, and processing.
Reusable	...ensuring that your data - and their related metadata - are well-described and indicate how they can be reused with appropriate licensing.

This means that prior to depositing your data, you need to package them together with a bit of metadata or contextual information. This information will include the context of your data within subject headings and keywords, the license that you will assign to the data and the added step of trying to ensure that your file types can be interoperable - open to being usable by others on as many different platforms as possible. This is a crucial step when depositing your data, as you will see throughout these guidelines, so make sure to take the time to complete the information about your data with as much relevant information as possible. Consider that as you share your data more and more people may come into contact with them and will therefore require a full scope of information about them:

---

[1] Force11. **Guiding principles for findable, accessible, interoperable and re-usable data publishing version b1.0**. Online. Available: https://www.force11.org/fairprinciples.
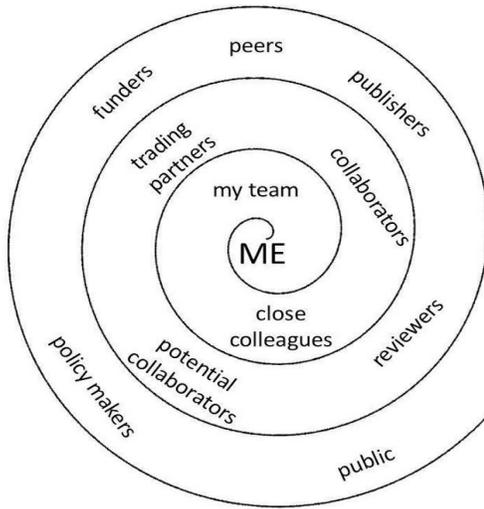
Slide made by Carol Goble: https://zenodo.org/record/2037060

Working on the FAIR data principles requires a bit of extra work, but this work is crucial in ensuring that your data do not get lost in the wild data pool where very few can find and reuse them. Within the ilifu consortium each of the participating institutions (Cape Peninsula University of Technology, Stellenbosch University, University of Cape Town and University of the Western Cape) has its own repository, meaning that researchers affiliated to these institutions should consider sharing and publishing their data on their own institutional repository if there is no domain-specific one in which they would rather deposit the data. Each of the ilifu institutional repositories is powered by Figshare for Institutions, a software solution provided by the software-as-a-service company Figshare. For more information contact the managers of your respective institutional research data repositories.

Repository name	Contact e-mail address
CPUT figshare	lockhartj@cput.ac.za
SUNScholarData	rdm@sun.ac.za
UWC figshare	library-research@uwc.ac.za
Zivahub	niklas.zimmer@uct.ac.za

2. How to log in

1. Go to the homepage of your institutional research data repository:
    1. Cape Peninsula University of Technology: https://cput.figshare.com/
    2. Stellenbosch University: https://scholardata.sun.ac.za/
    3. University of Cape Town: https://zivahub.uct.ac.za/
    4. University of the Western Cape: https://uwc.figshare.com/
2. Click on the red Log in button on the top right side of the screen
3. Use your institutional single sign on credentials to log in.
4. You will be taken to the SAFIRE single sign on screen the first time you log in.
5. You will need to press the "Yes, I agree" button to continue to your profile on the institution's research data repository.

3. How to create your user profile

The purpose of these guidelines is to assist you to **create your user profile**, and **associate the data you publish** with your **ORCiD profile**. Your user profile is a window to the world concerning who you are as a researcher, and the data associated with your research. It is important to **create a coherent public image** of who you are in order to **promote your work**, and **connect you** to other researchers with similar interests.

Profile picture

- We suggest that you use a recent image that portrays and identifies you professionally.

Professional details

- **ORCiD** identifier: Your user profile should be linked to your ORCiD identifier. For more information on how to sync your user account with ORCiD please see:
  How to link your user account with ORCiD.
  NB: **please sign into ORCID using your institutional account.**[2]
- First and last **name**: Include in your name all 'other names' you have published under, including abbreviations, middle initials and different surnames.
- Field of **interest**: Include your field of interest. If your field of interest is not listed, you can do a keyword search by typing it in the text box.
- **Location**:   Organisation name, location etc.

Social media links

- Facebook
- Twitter
- LinkedIn

---

[2] ORCiD Support. **Sign into ORCID using your institutional account.** Online. Available: https://support.orcid.org/hc/en-us/articles/360006894394-Sign-into-ORCID-using-your-institutional-account.

Biographical information

This section reveals who you are as a researcher. It details your research interests, history and career goals. Make use of this space to share your past experience and future prospects in an engaging way.

Publications

You may type in the references to your publications in this section. There is an **+Add publication** button that will allow you to scroll down and add as many publications as you like.

Once you are done, simply tick save changes, and go to view your public profile!

NB: We encourage you to **regularly update** your profile to ensure that it correctly represents your professional persona and your affiliations.

4. How to link your user account with ORCiD

What is ORCiD?

ORCiD is an open, non-profit, community-driven effort to create and maintain a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. For more information please go to: **https://orcid.org/content/about-orcid**. Figshare offers an ORCiD integration that, once connected, will allow you to push all of your public items from your institutional research data repository to ORCiD.

Connecting your user account to ORCiD

To push all of your public items from your institutional research data repository to ORCiD you need to ensure DataCite is authorised on your ORCiD account. DataCite is responsible for issuing the Digital Object Identifiers (DOIs) for data in your institutional research data repository. Once authorised, your published items will be displayed in the work section of your ORCiD profile. The following steps show you how to authorise DataCite on your ORCiD account:

1. **Sign in** at https://profiles.datacite.org/
2. Sign in with your **ORCiD**
3. **Authorise**

Sign in to your ORCiD account at: https://orcid.org/

1. Scroll down to the **Works** section and click **+Add works**
2. Select **Search & link**, you will be presented with the services you can link to
3. Select **DataCite**

Next, you will be presented with a permissions form. Upon reviewing the Privacy Policy, tick the box alongside **'Allow this permission until I revoke it'** and select **Authorise**.

You will then be taken to your **DataCite settings** page. To check that the authorisation has worked successfully head back to ORCiD and to your Account Settings tab. ORCiD should now be listed as a **Trusted organisation**.

Furthermore, you can **set up a DataCite profile** (if you do not already have one) and then give DataCite permission to automatically update your ORCiD record. You only have to do that once and DataCite will keep on updating your ORCiD record every time they mint a DOI, where your ORCiD iD is included in the metadata. See: **https://profiles.datacite.org/,** and click on ORCiD Auto-Update once you have a profile set up. In this way, you do not have to keep manually adding items to your ORCiD account. This is especially useful when you have a lot of data,

To then sync your ORCiD to your user profile, follow these steps:
1. Click **Connect** and **enable syncing** with your ORCiD
2. **Sign in** to ORCiD
3. **Authorise**

5. How to deposit a single file
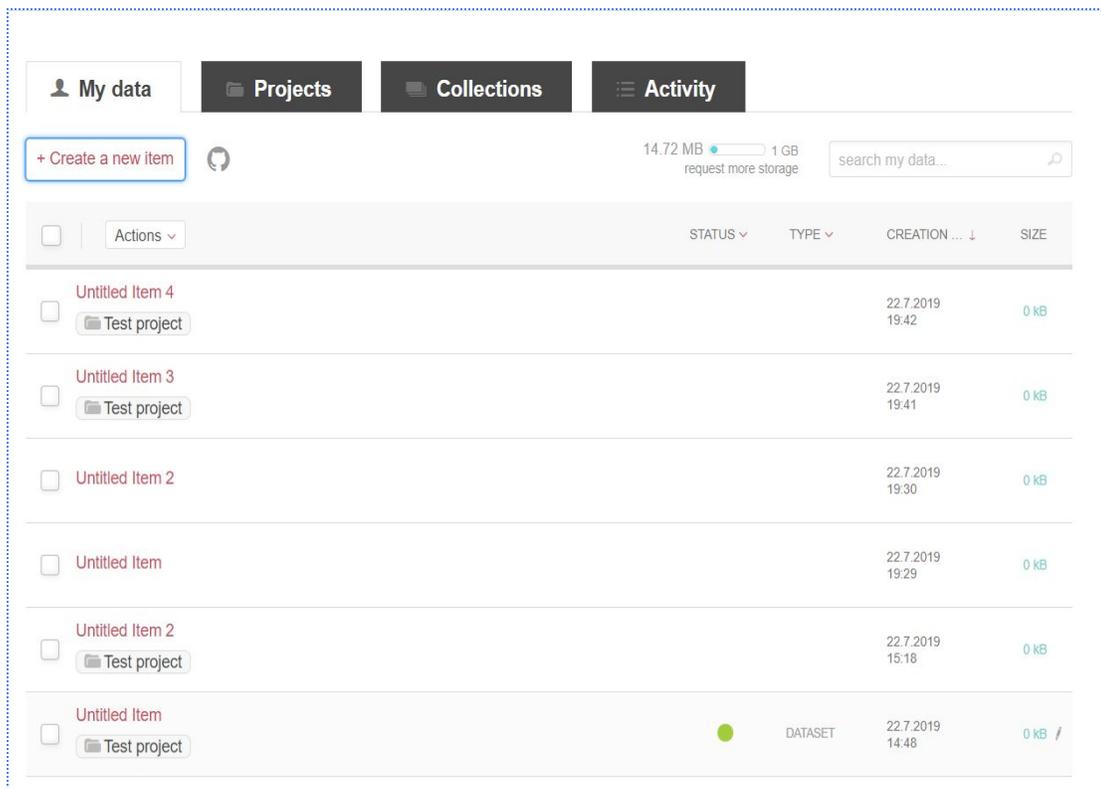
1. Log in to your institutional research data repository.
2. Make sure you are on the tab labelled '**My data**.'
3. Select the '**+ Create a new item**' button on the top left-hand side of the window or simply drag and drop your file on to the browser.

4. This will take you to the item's metadata submission form.

4.1. The metadata submission form can be saved, as is, at this point, for later completion, by clicking the **Save Changes** button at the bottom of the form. It is strongly suggested that you give the item a meaningful title at this point. The item can be edited before being published.

4.2. The fields with a green circle ● next to them are mandatory fields which must be completed before the item can be published. The unmarked fields are optional, but should be completed where appropriate.

4.3. See section <u>10. Completing the metadata submission form</u> for assistance on completing the form if needed.

4.4. Each time you have finished editing a metadata submission form click the **Save changes** button.

5. Click on **Save Changes** to complete the form.

6. Once you have saved the item you will be directed back to the **My data** tab. The newly created item (**Single Sample Spreadsheet**) will now appear in your list of items.

6. How to edit an item

1. Log in to your institutional research data repository.

2. Make sure you are on the tab labelled '**My data**.'

3. Click on the item which you wish to edit. In the example which follows the file titled **Untitled Item** (fifth from the top) will be edited.

4. This will take you to the item's metadata submission form.

to upload, drag file(s) on the page or browse

Metadata record only   Link file

needed to publish ●

Title   **Tips**

Untitled Item   ●   Use this form to edit all
                                              information related to your data.
Authors   Please be as descriptive as
                                          ●   possible. The file upload is
figshare admin sun ✕   independent from the rest of the
                                              form, so you don't need to save
Search co-authors by name, full email or ORCID. Hit enter after each.   an upload. This message will be
                                              replaced with helpful tips and
Categories   suggestions as you begin
                                              interacting with the form.
Select categories   ▼ ●

Item type   ●

Select item type   ▼ ●   Preview item (private)
   Edit timeline
Keyword(s)   Last edited on 22 Jul 2019 -
                                          ●   19:29 by Samuel Simango
Add keywords for easy discovery. Hit enter after each

Description

Describe your data as well as you can. Formatting is preserved   ●
when pasting from other sources and counts towards character
limits

B *I* U̲ X₂ X²

Funding

Search grant by name/number or add your own

Cancel   DOI   🔗   🚫   ☐ Publish   Save changes

5. Edit the metadata submission form. In the example below the title (**Untitled Item**) has been deleted and replaced with a new one (**File with a new title**).

6. Click on **Save Changes** to complete the form.

7. Once you have saved the item you will be directed back to the **My data** tab. The newly edited title will now reflect in your list of items.

8. If you make changes to other metadata fields and would like to review them click on the item on the list in order to see a preview of the item. Edit the item again if changes are still required.

    NB: If you are editing an item that has already been published, any edits you make from **My Data** will only reflect on the published copy once you have re-published the item. This may cause a version change, depending on which fields are edited.

7. How to deposit multiple files

1. Log in to your institutional research data repository.
2. Make sure you are on the tab labelled '**My data**.'
3. Select the **+ Create a new item** button from the top left of the window or simply drag and drop your files on to the browser.

4. Dragging more than one file at a time onto the browser will prompt you to create one item with multiple files or create individual new items.
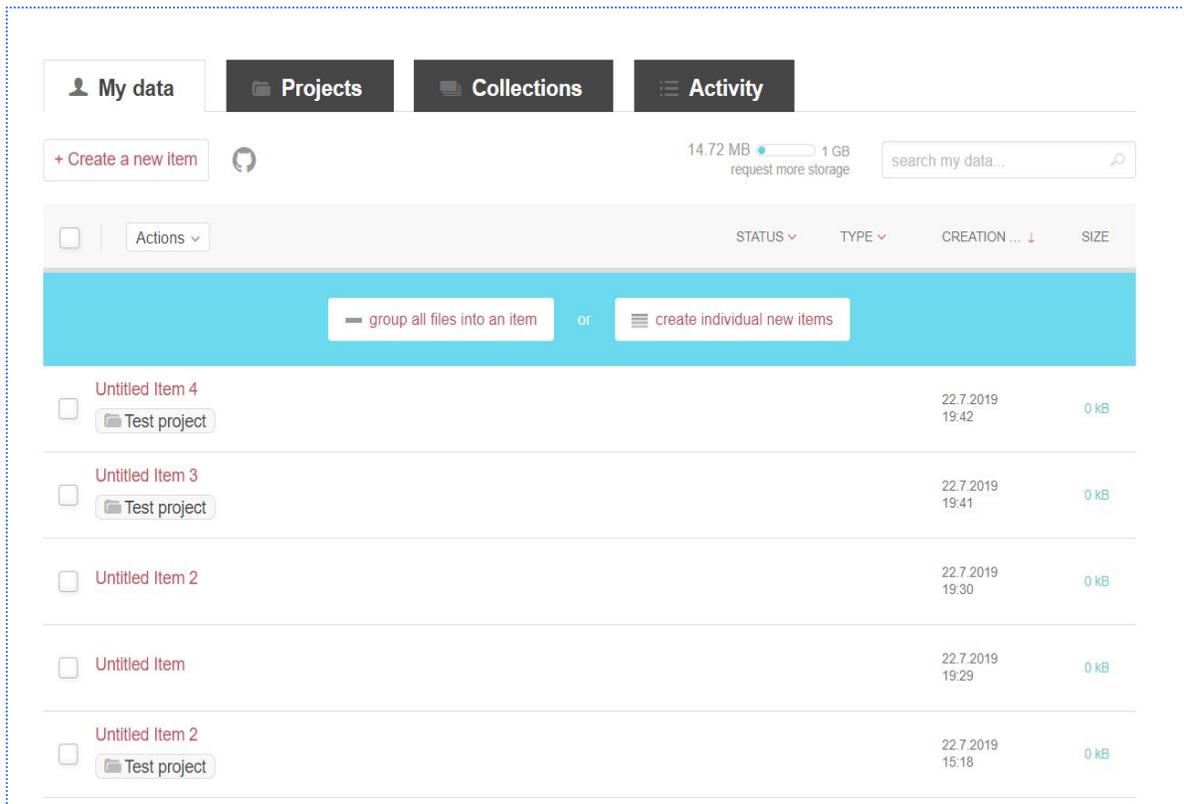
5. Select your preferred option by clicking on either the **group all files into an item** or **create individual new items** button (in this example we have opted to create individual new items).

6.   This will take you to a metadata submission form, which only displays one of the newly created items at the top.

6.1.   A metadata submission form has been opened for each item.

6.2.   The metadata submission forms can be saved at this point, as they are, for completion later by clicking the **Save Changes** at the bottom of the form. As each is saved the next will appear.

6.3.   The fields with green circles ● next to them are mandatory fields which must be completed before the item can be published. The fields without any green circles are optional, but should be completed if possible.

6.4.   See section 10. Completing the metadata submission form for assistance on completing the form if needed.

6.5.   Each time you have finished editing a metadata submission form click the **Save changes** button.
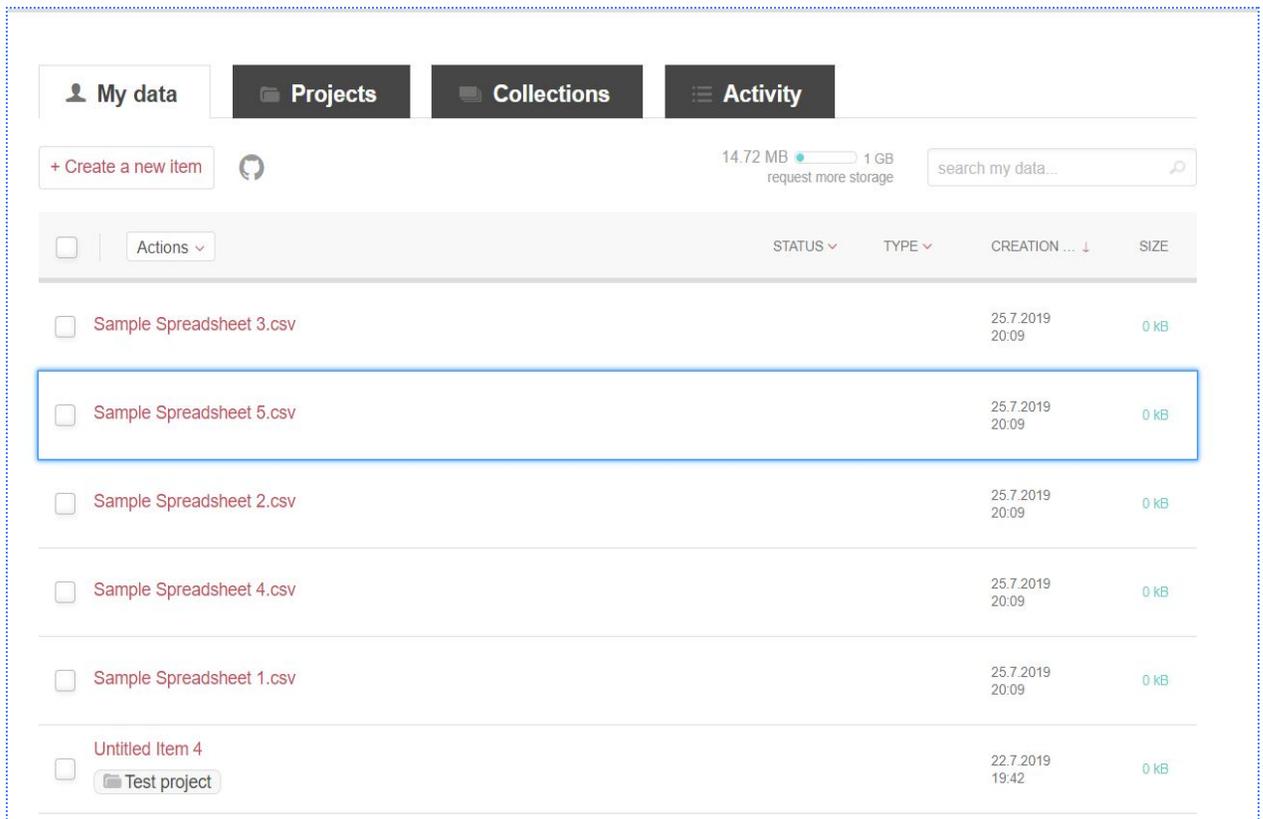
7.   Once you have saved each item you will be taken back to your **My data** page. You will notice that each of the newly created items (**Sample Spreadsheets 1 - 5**) now appears in your list of items.

8. How to edit a batch of files

NB: Incorrect batch edits cannot easily be undone and must be corrected for each item separately. Some can be overwritten in batch, but one must use this tool cautiously. It is highly recommended that batch editing first be tested by editing a single item and then previewing the item in order to ensure that the desired result has been achieved before applying an edit in batch to many items.

1. Log in to your institutional research data repository.
2. Make sure you are on the tab labelled '**My data**.'

3. Select the files which you would like to edit by ticking the check boxes on the left hand side of the page.
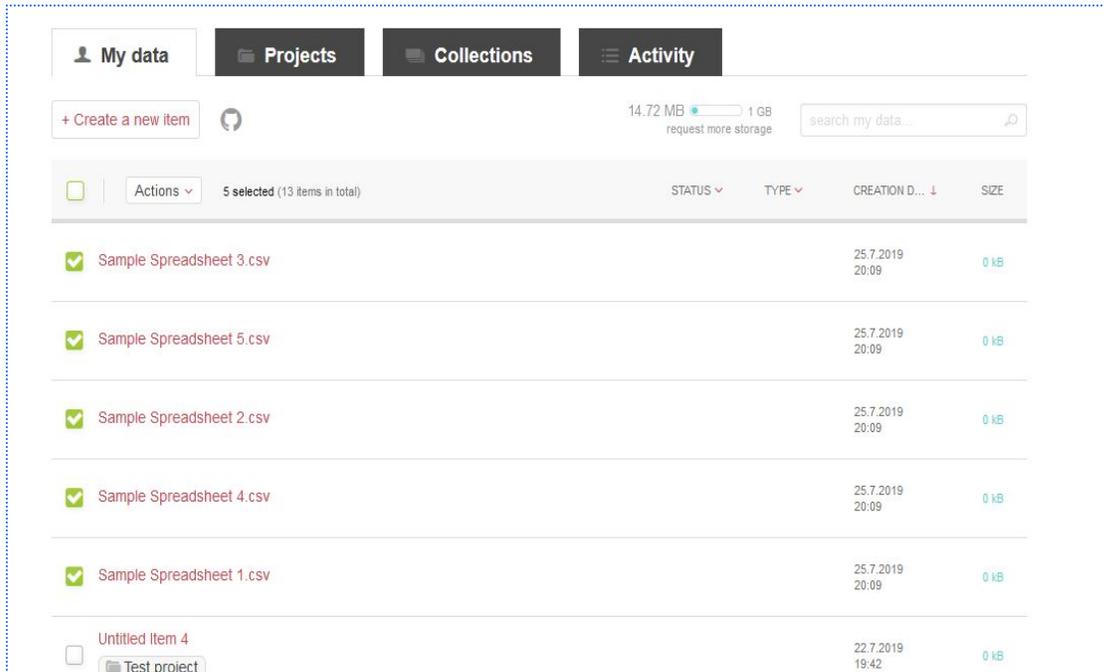
4. Click on **Actions** at the top of the page, and select **Edit in batch**.
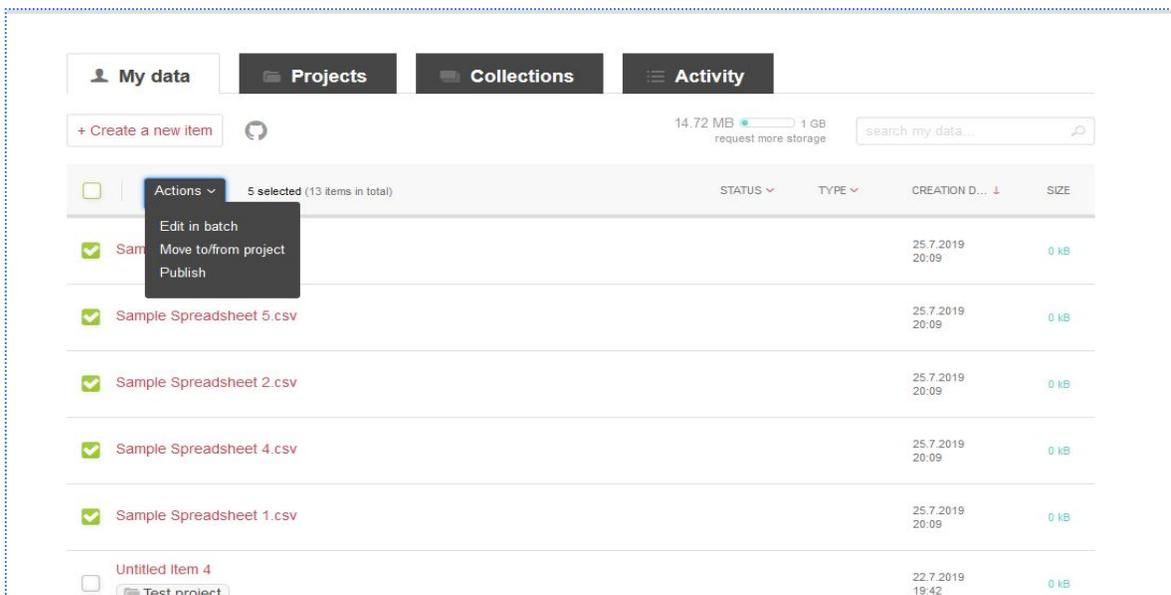
5. Once you have selected the items to be edited a new window will open with a list of the metadata fields that can be edited in batch.

6. Select the metadata field which you wish to edit and make the desired changes (in this example the keyword sample has been added).
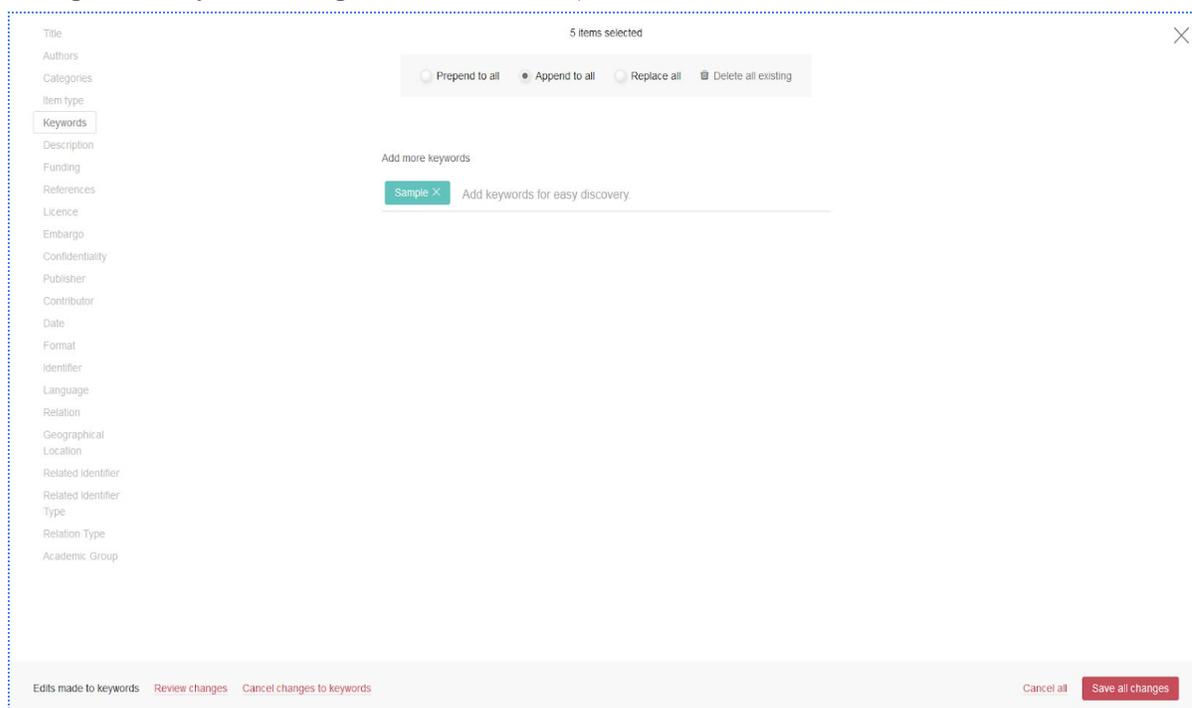
7. After you have finished editing the metadata field save the changes by clicking on the **Save all changes** button.

8. You will be prompted to review the changes made, but this will not show you a preview of the published item. This will enable you to go back and edit your changes by clicking the **back to editing** link or to confirm your changes by clicking the **Confirm and save all** button.
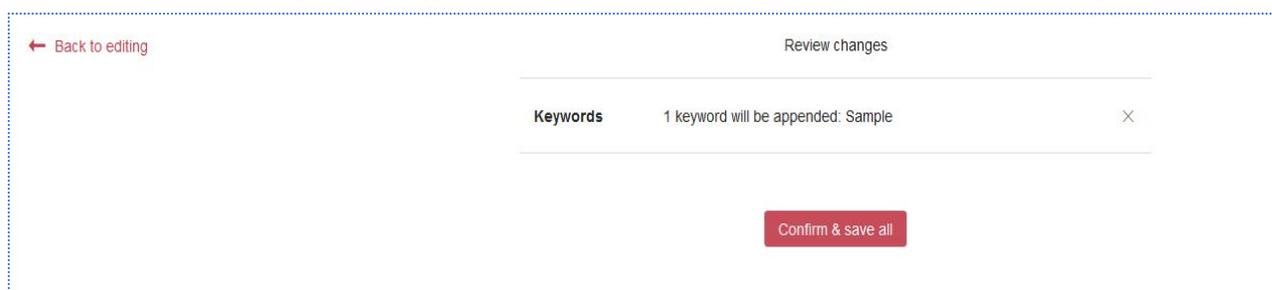
9. The changes will be saved and updated. Click done in order to complete the process and return back to the **My data** page.
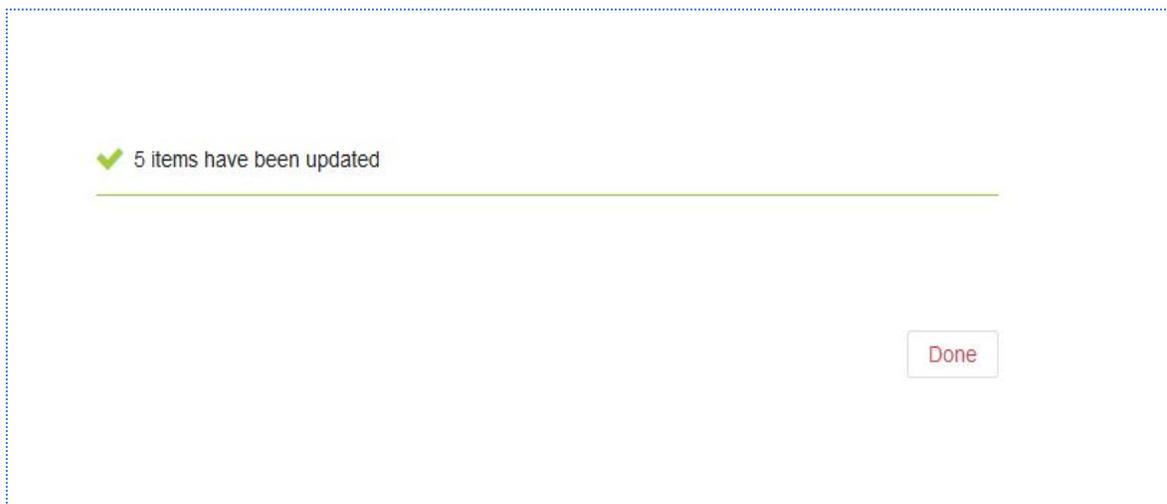
10. Once you have returned to My data you can preview any of the files in order to check that the changes reflect accurately.

11. In this example the term **sample** was successfully added to the keyword metadata field.

9. How to publish items

1. Log in to your institutional research data repository.
2. Make sure you are on the tab labelled '**My data**.'

3. Select the item which you wish to publish. In the example which follows the file titled **Weather Observations_Stellenbosch University Library_V5** will be submitted for publication.

4. This will take you to the item's metadata submission form. All fields with a green circles ●
next to them are mandatory fields and must be completed before the item can be published.
The fields without any green circles should also be completed if applicable. See section 10.
Completing the metadata submission form for assistance if needed.

5. After you have completed the mandatory fields in the metadata submission form click **Save
Changes**.

6. Select the **Publish** tick box. This will change the **Save Changes** button to the **Publish Item** button.

7. Click the **Publish Item** button to start the publishing process.

8. A banner will appear and prompt you to confirm that you wish to publish. **Importantly** it will also ask you to confirm which **open license** you wish to publish under.

9. Click the **terms** link. A new tab will open up, which contains the terms for your institutional research data repository.
10. **Make sure that you read and understand the terms to which you agree before publishing any items.**

11. After reading the terms and deciding that all is correct click the **Yes, publish** button.

12. The published item will be submitted to your institutional research data repository's review team. When the submission for publication has been accepted the status of the file will change to show a solid green dot.

13. If you would like to view the file in the public interface you may do so by clicking on your university's logo which is located on the top-left hand corner of the page. The institutional research data repository homepage will display the most recently published items. If your item does not appear here simply search for it.

10. Completing the metadata submission form

The metadata submission form is your best opportunity to pass on any and all relevant information about your data to future researchers and so this guide will briefly touch on all of the fields on the record with suggested or recommended goals and purposes for each field. Be aware that only the fields marked with a green dot are **required**, but an attempt should be made to complete all fields that may be relevant.

1. Title

Your institutional research data repository will automatically name your item with the filename of the first item you uploaded, but these file names are often cryptic and so it is highly recommended that you change the title.

Some tips for title creation:
- Try to be **concise**, be specific and try to use terms that would be recognised in the field of study to which the respective research data are related.
- The title should **uniquely identify** this specific data set and not be a generic reference to a type of data.
- Try to **avoid generic terms**, especially if your title is getting too long. Terms like "A spreadsheet of…" or "raw data" can be used to better explain the data set in the description.
- The description and title fields potentially have the most influence on a human reader in terms of item value for reuse and so it is important that some care is taken to try and maximise the impact of these fields.

2. Authors

The authors field should contain all those authors who deserve credit for the creation of the data set.
- The uploading account holder will always be added as the first author.
  - If you are not an author and are simply the person publishing the data set, be sure to remove your name from the list.
  - If you are an author be sure your name appears in the correct ordered place.
  - Be aware you can click and drag the author names to reorder them.
- All authors should be **strongly** advised to have an account on their institutional research data repository.

- ○ This will allow their profiles to automatically be updated with statistics tracked by the institutional research data repository, such as views, downloads and citations.
- ○ This will also allow their profiles to be linked to their ORICID, further automating the collection of metrics on their academic profiles.

3. Categories

At least one category needs to be selected for the generation of a DOI. As the category serves as an aggregator that will make the data set more findable it is useful to try and add all categories or research types that could find the data useful, not just the category of research that produced the data set.

- ● Select the category that most closely matches the research field that produced the data first as this is the category used to create the DOI.
- ● Select any other research category that may be applicable, but try not to generalise the items usefulness too much. One to five categories should be sufficient.

4. Item Type

This is a drop down list of fairly self-explanatory options. One option to take note of is "Data set" which is indexed for searches by Google Dataset Search. This adds greatly to the findability of your item, but this option should only be used if it is suitable for the item being published.

5. Keywords

Keywords can be a bit of a mystery but their ultimate goal is to assist the findability of the data set. Keywords can be seen in two different ways, either as specific search terms that could/should be used to discover this particular item, or as aggregation terms that could be used to classify the data into a research domain.

- ● Keywords do not have to be single words and can be phrases or sentences.
- ● All of the item record is searchable so terms used elsewhere may not need to be repeated to increase the findability of the item. For example, if the Category chosen defines the field of study it may not be necessary to repeat it as a keyword.
- ● We suggest "tiered" keywords
    - ○ Most keywords should be specifically focussed on the findability of this specific item.

- ○ The next level of keywords should try to class the item into its research domains or disciplines.
- ○ If applicable, add a keyword or two for the research group that is publishing the data so that all of their publications on the institutional research data repository can be aggregated with a single search term. This is not necessary for listed authors.
- ○ Add a keyword for your institution. Your institutional research data repository can be searched from figshare.com so it is best not to assume that all searchers will be aware of the institutional research data repository's existence.
- ○ Be careful not to "keyword bomb", ten to fifteen keywords are sufficient. While it is important that your item is findable it should only appear for relevant searches.

6. Description

This is the "human readable" part of the item record and care should be taken to describe the data set sufficiently to allow other researchers to evaluate its usefulness or applicability fairly easily. The data set is likely to be viewed by a wide variety of potential users and so some care should be taken to describe the item to a 'non-expert'. Having stated that it is likely though that most users who have found the data set will have some expert knowledge of the subject matter. For this reason we recommend that:

- The description is written with the most pertinent, expert level information at the beginning. This will allow an expert user to evaluate it quickly.
- Additional information or explanations should then follow with the description becoming less specific and more generic towards the end. Try to write it in such a way that an expert user can get the information they need quickly, but someone with less expertise would benefit from reading the full description.
- The description and title fields potentially have the most influence on a human reader in terms of item value for reuse and so it is important that some care is taken to try and maximise the impact of these fields.

The text box where you place your description is not very user friendly and so here are some tips:

- There is a character limit to the description and markup language will add to that character count. It is recommended to not paste text into this box unless the formatting

has been removed. One easy way to do this is to paste into Notepad first to remove formatting and then copy and paste into the description from there.

- Formatting applied to text once it is in the description does not add to the character count.
- Copying and pasting from other descriptions incurs no penalty so if you have a description element that repeats, such as a description of a project or research unit, this can be set up for one item and then copied and applied to related items, either singly or using the batch edit tool.

7. Funding

Any relevant funders can be entered here, particularly if you are publishing your data as part of a funder requirement. Be aware that multiple funders can be added, each in their own field.

8. References

Here you can list URLs and DOIs that have significance to the item being published.

- URLs could link to the institution, research group, funder details, etc.
- DOIs should link back to publications that reference the item, for example if this is a data set being published to compliment a paper or journal article then the DOI for that publication should be listed in the references.
- References can be added and removed from an item **after publication** without triggering a version change.

9. Licensing

The decision of which license to publish under can be complex. The following questions should help to determine which open license to use:

- Was the data set funded? A funder may stipulate the open license that should be used, especially if the publication of the data is a funder requirement.
- Does your institution have an Open Access policy? Do they specify the use of a particular license?
- Is the item being published software or code? Many of the licenses available on your institutional research data repository relate specifically to software and code (Open Source).
- Should any restrictions be placed on the re-use of the item being published? Use restrictions should help to determine the open license to be used.

A guide to some of the open licenses that may be available on your institutional research data repository can be found below. Choose the license that best suits the restrictions of your data set. If none of the available licenses are applicable then consider:

- Asking your institutional research data repository management team to add a suitable license if one can be identified.
- Whether the item or data set should be published openly.

10. Choosing a License

As all of the institutional research data repositories referred to in these guidelines are open, any item published on these repositories will have to be published under an open license. By making your research data publicly available, you retain ownership of your research objects (as is often not the case with traditional publishing). With the FAIR data principles in mind, data should be as open as possible, and as closed as necessary.

Figshare offers a few different options for choosing a license for research data, and ilifu's individual partner institution repositories have different licensing options available to researchers. Please note that the availability of specific licenses at the different ilifu partner institutions will vary. Below are the possible licenses which could be available for selection at your institution.

10.1 CC-0

This is the most open Creative Commons license. There is no copyright reserved in a CC-0 license, and users are not obliged to cite the source of the data set.[3]

10.2 CC-BY 4.0 (Attribution only)

This license lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation,in the form of a citation, should they use or refer to the research object.[4] This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

---

[3] Creative Commons. **Public Domain License**. Online. Available: https://creativecommons.org/share-your-work/public-domain/cc0/.

[4] Creative Commons. **Attribution License**. Online. Available: https://creativecommons.org/licenses/by/4.0/.

10.3   GPL (GNU General Public License)

The GPL licenses (versions 1, 2 and 3) are a series of widely used free software licenses that guarantee end users the freedom to run, study, share, and modify the software. They are all copyleft licenses, which means that any derivative work must be distributed under the same or equivalent license terms. GPL was the first copyleft license for general use, and has historically been one of the most popular software licenses in the free and open-source software domain.[5]

10.4   The MIT License

The MIT license is a permissive license placing very limited restriction on reuse. It has high license compatibility because it can be re-licensed under other licenses.[6] The MIT license is compatible with many copyleft licenses, such as the GPL.

10.5   Apache 2.0

Another permissive and free software license, the Apache license is written by the Apache Software Foundation (ASF) and allows users to use the software for any purpose, to distribute it, to modify it, and to distribute modified versions of the software under the terms of the license, without concern for royalties.[7] The Apache License 2.0 is compatible with the GPL 3.0.

10.6   CC-BY-SA (Attribution-ShareAlike)

This license lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms.[8] This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

---

[5] GNU Operating Software. **GNU General Public License**. Online. Available: https://www.gnu.org/licenses/gpl-3.0.html.
[6] Choose an Open Source License. **MIT License**. Online. Available: https://choosealicense.com/licenses/mit/.
[7] Apache Software Foundation. **Apache 2.0 License**. Online. Available: https://www.apache.org/licenses/.
[8] Creative Commons. **Attribution-ShareAlike License**. Online. Available: https://creativecommons.org/licenses/by-sa/4.0/.

10.7   CC-BY-NC (Attribution-NonCommercial)

This license lets others remix, adapt, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they do not have to license their derivative works on the same terms.[9]

10.8   ODC-BY (Open Data Commons Attribution License)

This license relates to databases. The license itself permits users to share databases, create works from databases and to adapt databases provided that any public use of a database, or works produced from a database have been attributed to the database's original creators.[10]

11. Apply Embargo

An embargo simply allows you to 'pre-publish' the item and set the date for when the item becomes publicly available.

12. Make Files Confidential

Making files confidential allows items to be published, but only the associated metadata will be made visible. The uploaded files will not be accessible to users of the institutional research data repository. This option is generally used in the case of confidential data where it is important that the existence of the data be made known, but the items themselves are not suitable for open publishing. Access to the item is arranged between a prospective researcher and the item publisher or owner.

13. Generate Private Link

Generating a private link allows for the sharing of the item before it is published. This may be necessary inside a research group or simply to allow for the item to be previewed by other parties before it is published.

---

[9] Creative Commons. **Attribution-NonCommercial License**. Online. Available: https://creativecommons.org/licenses/by-nc/4.0/.

[10] Open Data Commons. **ODC Attribution License**. Online. Available: https://opendatacommons.org/licenses/by/summary/index.html.

14. Reserve Digital Object Identifier

Reserving a DOI generates the items DOI before publication. This allows other publications to reference the item correctly before it is published. This is generally used in the case of an article and a data set being published together - in which case it is necessary to have the data set DOI for the article's references. Every time your data is published on a repository, your data is assigned a DOI which ensures that there is one persistent and unbroken link to your data. This will ensure that as your data gets downloaded, reused, and cited, you will be able to track the usage.